PCT/GB 2003 / 003686

INVESTOR IN PEOPLE

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 25 SEP 2003
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Stephen Hordley

Dated 18 June 2003

THIS PAGE IS BLANK

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
G
23 DEC 2002
LONDON

The Patent Office

1/77

27DEC02 E773396-1 D02882
P01/7700 0.00-0230037.4

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

23 DEC 2002

The Patent Office

Concept House
Cardiff Road
Newport
South Wales NP10 8QQ

1.	Your reference	**CDM/DCM/61500**
2.	Patent application number *(The Patent Office will fill in this part)*	0230037.4
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	**SOLEXA LIMITED** **Chesterford Research Park** **Little Chesterford** **Nr. Saffron Walden** **Essex CB10 1XL**
	Patents ADP number *(if you know it)*	07533698002
	If the applicant is a corporate body, give the country/state of its incorporation	**GB**
4.	Title of the invention	**Modified Nucleotides**
5.	Name of your agent *(if you have one)*	**BOULT WADE TENNANT**
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	**VERULAM GARDENS** **70 GRAY'S INN ROAD** **LONDON WC1X 8BT**
	Patents ADP number *(if you know it)*	**42001**
6.	If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country / Priority application number (if you know it) / Date of filing *(day/month/year)*
7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application / Date of filing *(day / month / year)*
8.	Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if: a) any applicant named in part 3 is not an inventor, or b) there is an inventor who is not named as an applicant, or c) any named applicant is a corporate body. See note (d))*	**Yes**

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	25
Claim(s)	6
Abstract	
Drawing(s)	4 + 4

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(Please specify)*	

11 I/We request the grant of a patent on the basis of this application.

Signature: Boult Wade Tennant

Date: 20 December 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

Colm Murphy
020 7430 7500

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents ct 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

Modified Nucleotides

FIELD OF THE INVENTION

The invention relates to modified nucleotides. In particular, this invention discloses nucleotides having a removable blocking group and their use in polynucleotide sequencing methods.

BACKGROUND

Advances in the study of molecules have been led, in part, by improvement in technologies used to characterise the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for sequence analysis and the study of hybridisation events.

An example of the technologies that have improved the study of nucleic acids, is the development of fabricated arrays of immobilised nucleic acids. These arrays consist typically of a high-density matrix of polynucleotides immobilised onto a solid support material. See, e.g., Fodor et al., Trends Biotech. 12:19-26, 1994, which describes ways of assembling the nucleic acids using a chemically sensitized glass surface protected by a mask, but exposed at defined areas to allow attachment of suitably modified nucleotide phosphoramidites. Fabricated arrays can also be manufactured by the technique of "spotting" known polynucleotides onto a solid support at predetermined positions (e.g., Stimpson et al., Proc. Natl. Acad. Sci. USA 92:6379-6383, 1995).

Sequencing by synthesis of DNA ideally requires the controlled (i.e. one at a time) incorporation of the

correct complementary nucleotide opposite the oligonucleotide being sequenced. This allows for accurate sequencing by adding nucleotides in multiple cycles as each nucleotide residue is sequenced one at a time, thus preventing an uncontrolled series of incorporations occurring. The incorporated nucleotide is read using an appropriate label attached thereto before removal of the label moiety and the subsequent next round of sequencing. In order to ensure only a single incorporation occurs, a reversible structural modification ("blocking group") of the sequencing nucleotides is required to ensure a single nucleotide incorporation but which then prevents any further nucleotide incorporation into the polynucleotide chain. The blocking group must then be removable, under reaction conditions which do not interfere with the integrity of the DNA being sequenced. The sequencing cycle can then continue with the incorporation of the next blocked, labelled nucleotide. In order to be of practical use, the entire process should consist of high yielding, highly specific chemical and enzymatic steps to facilitate multiple cycles of sequencing.

SUMMARY OF THE INVENTION

In the present invention, a nucleoside or nucleotide molecule comprises a purine or pyrimidine base and a ribose or deoxyribose sugar moiety which has a removable 3'-OH blocking group covalently attached thereto, which renders the molecules useful in techniques requiring blocking of the 3'-OH group to prevent incorporation of additional nucleotides, such as for example in sequencing reactions, polynucleotide synthesis, nucleic acid amplification, nucleic acid hybridisation assays, single nucleotide polymorphism studies, and other such techniques.

Therefore, according to a first aspect of the invention there is provided a modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-OH blocking group covalently attached thereto, such that the 3' carbon atom has attached a group of the structure

$$-O-X$$

wherein X is any of $-C(R')_2-O-R$, $-C(R')_2-N(R)_2$,$-C(R')_2-N(H)R$, $-C(R')_2-S-R$ and $-C(R')_2-F$,

wherein each R group is or is part of a removable protecting group;

each R' is independently H or an alkyl; and wherein said molecule may be reacted to yield an intermediate in which each R group is exchanged for H or, where X is $-C(R')_2-F$, the F is exchanged for OH, SH or NH_2, preferably OH, which intermediate dissociates under aqueous conditions to afford a molecule with a free 3'OH;

with the proviso that where X is $-C(R')_2-S-R$, both R' groups are not H.

The intermediates produced advantageously spontaneously dissociate under aqueous conditions back to the natural 3' hydroxy structure, which permits further incorporation of another nucleotide. Any appropriate protecting group may be used, as discussed hereinafter. Preferably, X is of formula $-C(R')_2-O-R$, $-C(R')_2-N(R)_2$, $-C(R')_2-N(H)R$ and $-C(R')_2-SR$. Particularly preferably, X is of the formula $-C(R')_2-O-R$, $-C(R')_2-N(R)_2$, and $-C(R')_2-SR$. One example of groups of structure -O-X wherein X is $-C(R')_2-N(R)_2$ are those in which $N(R)_2$ is azido ($-N_3$). One preferred such example is azidomethyl wherein each R' is H. R may also be a benzyl group or a substituted benzyl group in an alternative embodiment.

The molecule can be linked via the base to a detectable label by a desirable linker, which label may be a fluorophore, for example. The linker can be acid labile, photolabile or contain a disulphide linkage. Other linkers may be employed in this invention as described in greater detail below.

The invention also features a method of controlling the incorporation of a nucleotide molecule complementary to the nucleotide in a target single stranded polynucleotide in a synthesis or sequencing reaction comprising incorporating into the growing complementary polynucleotide a molecule according to the invention, the incorporation of said molecule preventing or blocking introduction of subsequent nucleoside or nucleotide molecules into said growing complementary polynucleotide. The incorporation of the molecule may be accomplished via a terminal transferase, a polymerase or a reverse transcriptase.

Additionally, the invention also comprises a method for determining the sequence of a target single stranded polynucleotide comprising monitoring the sequential incorporation of complementary nucleotides, wherein the nucleotide is a molecule according to the invention as hereinbefore described that is linked to a detectable label via a cleavable linker and wherein the identity of each nucleotide molecule incorporated is determined by detection of the label linked to the base, and said blocking group and said label is removed prior to introduction of the next complementary nucleotide. In a preferred embodiment of this aspect of the invention, the blocking group and the label may be removed in a single chemical treatment step. Thus, in a preferred embodiment of the invention, the blocking group is chemically identical on each of the bases A,T,C and G and the block is cleaved simultaneously with the label.

Furthermore, preferably the blocked and labelled modified nucleotide constructs of the nucleotide bases A,T,C and G are recognised as substrates by the same polymerase enzyme.

In another aspect, the invention provides a method for determining the sequence of a target single-stranded polynucleotide, comprising:

(a) providing modified nucleotide molecules according to the hereinbefore described invention, linked by the base to a detectable label via a cleavable linker and wherein the detectable label linked to each type of nucleotide can be distinguished upon detection from the detectable label used for other types of nucleotides;

(b) incorporating the nucleotide into the complement of the target single stranded polynucleotide;

(c) detecting the label of the nucleotide of (b), thereby determining the type of nucleotide incorporated;

(d) removing the label of the nucleotide of (b) and the blocking group;

(e) optionally repeating steps (b)-(d) one or more times;

thereby determining the sequence of a target single-stranded polynucleotide. In this aspect, the label of the nucleotide and the blocking group are preferably removed in a single chemical treatment step. Thus, in a preferred embodiment of the invention, the blocking group is chemically identical on each of the bases A,T,C and G and the block is cleaved simultaneously with the label. Furthermore, preferably the blocked and labelled modified nucleotide constructs of the nucleotide bases A,T,C and G are recognised as substrates by the same polymerase enzyme.

In the methods described herein, each of the nucleotides can be brought into contact with the target

sequentially, with removal of non-incorporated nucleotides prior to addition of the next nucleotide, where detection and removal of the label and the blocking group is carried out either after addition of each nucleotide, or after addition of all four nucleotides.

In the methods, all of the nucleotides can be brought into contact with the target simultaneously, i.e., a composition comprising all of the different nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and the blocking group.

The methods can comprise a first step and a second step, where in the first step, a first composition comprising two of the four types of modified nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where in the second step, a second composition comprising the two nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

The methods described herein can also comprise a first step and a second step, where in the first step, a composition comprising one of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where in the second step, a second

composition comprising the three nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

The methods described herein can also comprise a first step and a second step, where in the first step, a first composition comprising three of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and where in the second step, a composition comprising the nucleotide not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and where the first steps and the second step can be optionally repeated one or more times.

In a further aspect, the invention features a kit, where the kit includes: (a) individual nucleotides according to the hereinbefore described invention, where each nucleotide preferably has a base that is linked to a detectable label via a cleavable linker, and where the detectable label linked to each nucleotide can be distinguished upon detection from the detectable label used for other three nucleotides; and (b) packaging materials therefor. The kit can further include an enzyme for incorporating the nucleotide into the complementary nucleotide chain and buffers appropriate for the action of the enzyme in addition to appropriate chemicals for removal of the blocking group and the

detectable label, which can preferably be removed by the same chemical treatment step.

The nucleotides/nucleosides are suitable for use in many different DNA-based methodologies, including DNA synthesis and DNA sequencing protocols, including those involving single molecular arrays, particularly those disclosed in International Patent Application Publication No. WO00/06770.

According to another aspect of the invention, a method for determining the sequence of a target polynucleotide comprises monitoring the sequential incorporation of complementary modified nucleotides according to the invention, wherein the nucleotides comprise a detectable label linked to the base portion of the nucleotide via a cleavable linker, incorporation being detected by monitoring the label, and the label and blocking groups are removed to permit further nucleotide incorporation to occur.

DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic illustration of the blocking groups that can be used according to the invention.

Fig. 2 shows examplary nucleotide structures useful in the invention. For each structure, X can be H, phosphate, diphosphate or triphosphate. R_1 and R_2 can be the same or different, and can be selected from H, OH, or any group which can be transformed into an OH, including, but not limited to, a carbonyl. Some suitable functional groups for R_1 and R_2 include the structures shown in Fig. 4.

Fig. 3 shows structures of linkers useful in the invention , including (1) disulfide linkers and acid

labile linkers, (2) dialkoxybenzyl linkers, (3) Sieber linkers, (4) indole linkers and (5) *t*-butyl Sieber linkers.

Fig. 4 shows some functional molecules useful in the invention, including some cleavable linkers and some suitable hydroxyl protecting groups. In these structures, R_1 and R_2 may be the same of different, and can be H, OH, or any group which can be transformed into an OH group, including a carbonyl. R_3 represents one or more substituents independently selected from alkyl, alkoxyl, amino or halogen groups. Alternatively, cleavable linkers may be constructed from any labile functionality used on the 3′-block. R_4 and R_5 can be H or alkyl, and R_6 can be alkyl, cycloalkyl, alkenyl, cycloalkenyl or benzyl. X can be H, phosphate, diphosphate or triphosphate.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to nucleotide or nucleoside molecules that are modified by the reversible covalent attachment of a 3' OH blocking group thereto, and which molecule may be used in reactions where blocked nucleotide or nucleoside molecules are required, such as in sequencing reactions, polynucleotide synthesis and the like.

Nucleotide, and more usually nucleotide triphospates, generally require a 3' OH blocking group so as to prevent the polymerase used to incorporate it into a polynucleotide chain, from continuing to replicate once the base on the nucleotide is added. There are many limitations on the suitability of a molecule as a blocking group. It must be such that it results in complete blockage of additional nucleotide molecules to the polynucleotide chain whilst

simultaneously being easily removable from the sugar moiety without causing damage to the polynucleotide chain. Furthermore, the modified nucleotide must be tolerated by the polymerase or other appropriate enzyme used to incorporate it into the polynucleotide chain. The ideal blocking group will therefore exhibit long term stability, efficient incorporation by the polymerase enzyme, total blocking of secondary or further incorporation and the ability to be removed under mild conditions that do not cause damage the polynucleotide structure, such as under aqueous conditions. These stringent requirements are formidable obstacles to the design and synthesis of the requisite modified nucleotides.

Reversible blocking groups for this purpose have been described previously but none of them generally meet the above criteria for DNA compatable chemistry.

Protecting groups which comprise the acetal functionality have been used previously as blocking groups. However, removal of such groups, e.g. ethers requires strongly acidic deprotections detrimental to the DNA molecule. The hydrolysis of an acetal however, results in the formation of an unstable hemiacetal intermediate which hydrolyses under aqueous conditions to the natural hydroxyl group. The inventors have utilised this concept and applied it further such that the invention resides in utilising blocking groups that include protecting groups to protect intermediate molecules that would normally hydrolyse under aqueous conditions. These protecting groups comprise a second functional group that stabilises the structure of the intermediate but which can be removed at a later stage following incorporation into the polynucleotide. Protecting groups have been used in organic synthesis reactions to temporarily mask the characteristic

chemistry of a functional group because it interferes with another reaction.

As is known in the art, a "nucleotide" consists of a nitrogenous base, a sugar, and one or more phosphate groups. They are monomeric units of a nucleic acid sequence. In RNA, the sugar is a ribose, and in DNA a deoxyribose, i.e. a sugar lacking a hydroxyl group that is present in ribose. The nitrogenous base is a derivative of purine or pyrimidine. The purines are adenine (A) and guanine (G), and the pyrimidines are cytosine (C) and thymine (T) (or in the context of RNA, uracil (U)). The C-1 atom of deoxyribose is bonded to N-1 of a pyrimidine or N-9 of a purine. A nucleotide is also a phosphate ester or a nucleoside, with esterification occurring on the hydroxyl group attached to C-5 of the sugar. Nucleotides are usually mono, di- or triphosphates.

A "nucleoside" is structurally similar to a nucleotide, but is missing the phosphate moieties. An example of a nucleoside analog would be one in which the label is linked to the base and there is no phosphate group attached to the sugar molecule.

Although the base is usually referred to as a purine or pyrimidine, the skilled person will appreciate that derivatives and analogs are available which do not alter the capability of the nucleotide or nucleoside to undergo Watson-Crick base pairing. "Derivative" or "analog" means a compound or molecule whose core structure is the same as, or closely resembles that of, a parent compound, but which has a chemical or physical modification, such as a different or additional side groups, which allows the derivative nucleotide or nucleoside to be linked to another molecule. For example, the base can be a deazapurine. The derivatives

should be capable of undergoing Watson-Crick pairing. "Derivative" and "analog" also mean a synthetic nucleotide or nucleoside derivative having modified base moieties and/or modified sugar moieties. Such derivaties and analogs are discussed in, e.g., Scheit, Nucleotide Analogs (John wiley & Son, 1980) and Uhlman et al., Chemical Reviews 90:543-584, 1990. Nucleotide analogs can also comprise modified phosphodiester linkages, including phosphorothioate, phosphorodithioate, alkylphosphonate, phosphoranilidate and phosphoramidate linkages. The analogs should be capable of undergoing Watson-Crick base pairing. "Derivative" and "analog", as used herein, may be used interchangeably, and are encompassed by the terms "nucleotide" and "nucleoside" as defined herein.

In the context of the present invention, the term "incorporating" means becoming part of a nucleic acid (eg DNA) molecule or oligonucleotide or primer. An oligonucleotide refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are formed by a phosphodiester bond between the 3' position of the pentose on one nucleotide and the 5' position of the pentose on an adjacent nucleotide.

The blocking groups used in accordance with the invention are shown schematically in Figure 1, and generally concern protected hemiacetals, protected hemiaminals and protected hemithioacetals as reversible blocking groups. In a preferred embodiment the 3'OH group is blocked with, for example, an azidomethyl group as a protected form of a hemiaminal or 3,4-dimethoxybenzyloxymethyl group as a protected form of a hemiacetal. However, other suitable protecting groups may well be used as disclosed in Green & Wuts, Protective Groups in Organic Synthesis, John wiley &

Sons.

The term "alkyl" covers straight chain, branched chain and cyclic chain alkyl groups. Unless the context indicates otherwise, the term "alkyl" refers to groups having 1 to 8 carbon atoms, and typically from 1 to 6 carbon atoms, for example from 1 to 4 carbon atoms. Examples of alkyl groups include methyl, ethyl, propyl, isopropyl, n-butyl, isobutyl, tert-butyl, n-pentyl, 2-pentyl, 3-pentyl, 2-methyl butyl, 3-methyl butyl, and n-hexyl and its isomers.

Examples of cycloalkyl groups are those having from 3 to 10 ring atoms, particular examples including those derived from cyclopropane, cyclobutane, cyclopentane, cyclohexane and cycloheptane, bicycloheptane and decalin.

The term amino refers to groups of type NR^1R^2, wherein R^1 and R^2 are independently selected from hydrogen, a C_{1-6} alkyl group (also referred to as C_{1-6} alkylamino or di-C_{1-6} alkylamino).

The term "halogen" as used herein includes fluorine, chlorine, bromine and iodine.

The nucleotide molecules of the present invention are suitable for use in many different methods where the detection of nucleotides is required.

DNA sequencing methods, such as those outlined in U.S. Pat. No. 5,302,509 can be carried out using the nucleotides.

The present invention can make use of conventional detectable labels. Detection can be carried out by any suitable method, including fluorescence spectroscopy or by other optical means. The preferred label is a

fluorophore, which, after absorption of energy, emits radiation at a defined wavelength. Many suitable fluorescent labels are known. For example, Welch et al. (Chem. Eur. J. 5(3):951-960, 1999) discloses dansyl-functionalised fluorescent moieties that can be used in the present invention. Zhu et al. (Cytometry 28:206-211, 1997) describes the use of the fluorescent labels Cy3 and Cy5, which can also be used in the present invention. Labels suitable for use are also disclosed in Prober et al. (Science 238:336-341, 1987); Connell et al. (BioTechniques 5(4):342-384, 1987), Ansorge et al. (Nucl. Acids Res. 15(11):4593-4602, 1987) and Smith et al. (Nature 321:674, 1986). Other commercially available fluorescent labels include, but are not limited to, fluorescein, rhodamine (including TMR, texas red and Rox), alexa, bodipy, acridine, coumarin, pyrene, benzanthracene and the cyanins.

Multiple labels can also be used in the invention. For example, bi-fluorophore FRET cassettes (Tet. Letts. 46:8867-8871, 2000) are well known in the art and can be utilised in the present invention. Multi-fluor dendrimeric systems (J. Amer. Chem. Soc. 123:8101-8108, 2001) can also be used.

Although fluorescent labels are preferred, other forms of detectable labels will be apparent as useful to those of ordinary skill. For example, microparticles, including quantum dots (Empodocles, et al., Nature 399:126-130, 1999), gold nanoparticles (Reichert et al., Anal. Chem. 72:6025-6029, 2000) and microbeads (Lacoste et al., Proc. Natl. Acad. Sci USA 97(17):9461-9466, 2000) can all be used.

Multi-component labels can also be used in the invention. A multi-component label is one which is dependent on the interaction with a further compound for detection. The most common multi-component label used

in biology is the biotin-streptavidin system. Biotin is used as the label attached to the nucleotide base. Streptavidin is then added separately to enable detection to occur. Other multi-component systems are available. For example, dinitrophenol has a commercially available fluorescent antibody that can be used for detection.

The invention will be further described with reference to nucleotides. However, unless indicated otherwise, the reference to nucleotides is also intended to be applicable to nucleosides. The invention will also be further described with reference to DNA, although the description will also be applicable to RNA, PNA, and other nucleic acids, unless otherwise indicated.

The modified nucleotides of the invention use a cleavable linker to attach the label to the nucleotide. The use of a cleavable linker ensures that the label can, if required, be removed after detection, avoiding any interfering signal with any labelled nucleotide incorporated subsequently.

Cleavable linkers are known in the art, and conventional chemistry can be applied to attach a linker to a nucleotide base and a label. The linker can be cleaved by any suitable method, including exposure to acids, bases, nucleophiles, electrophiles, radicals, metals, reducing or oxidising agents, light, temperature, enzymes etc. Suitable linkers can be adapted from standard chemical blocking groups, as disclosed in Greene & Wuts, Protective Groups in Organic Synthesis, John Wiley & Sons. Further suitable cleavable linkers used in solid-phase synthesis are disclosed in Guillier et al. (Chem. Rev. 100:2092-2157, 2000).

The use of the term "cleavable linker" is not meant to imply that the whole linker is required to be removed from the nucleotide base. The cleavage site can be located at a position on the linker that ensures that part of the linker remains attached to the nucleotide base after cleavage.

The linker can be attached at any position on the nucleotide base provided that Watson-Crick base pairing can still be carried out. In the context of purine bases, it is preferred if the linker is attached via the 7 position of the purine or the preferred deazapurine analogue, via an 8-modified purine, via an N-6 modified adenosine or an N-2 modified guanine. For pyrimidines, attachment is preferably via the 5 position on cytosine, thymidine or uracil and the N-4 position on cytosine. Suitable nucleotide structures are shown in Fig. 2. For each structure in Fig. 2, X can be H, phosphate, diphosphate or triphosphate. R_1 and R_2 can be the same or different, and can be selected from H, OH, or any group which can be transformed into an OH, including, but not limited to, a carbonyl. Some suitable functional groups for R_1 and R_2 include the structures shown in Fig. 4.

Suitable linkers are shown in Fig. 3 and include, but are not limited to, disulfide linkers (1), acid labile linkers (2, 3, 4 and 5; including dialkoxybenzyl linkers (e.g., 2), Sieber linkers (e.g., 3), indole linkers (e.g., 4), t-butyl Sieber linkers (e.g., 5)), electrophilically cleavable linkers, nucleophilically cleavable linkers, photocleavable linkers, cleavage under reductive conditions, oxidative conditions, cleavage via use of safety-catch linkers, and cleavage by elimination mechanisms.

A. Electrophilically cleaved linkers.

Electrophilically cleaved linkers are typically cleaved by protons and include cleavages sensitive to acids. Suitable linkers include the modified benzylic systems such as trityl, p-alkoxybenzyl esters and p-alkoxybenzyl amides. Other suitable linkers include tert-butyloxycarbonyl (Boc) groups and the acetal system.

The use of thiophilic metals, such as nickel, silver or mercury, in the cleavage of thioacetal or other sulphur-containing protecting groups can also be considered for the preparation of suitable linker molecules.

B. Nucleophilically cleaved linkers.

Nucleophilic cleavage is also a well recognised method in the preparation of linker molecules. Groups such as esters that are labile in water (i.e., can be cleaved simply at basic pH) and groups that are labile to non-aqueous nucleophiles, can be used. Fluoride ions can be used to cleave silicon-oxygen bonds in groups such as triisopropyl silane (TIPS) or t-butyldimethyl silane (TBDMS).

C. Photocleavable linkers.

Photocleavable linkers have been used widely in carbohydrate chemistry. It is preferable that the light required to activate cleavage does not affect the other components of the modified nucleotides. For example, if a fluorophore is used as the label, it is preferable if this absorbs light of a different wavelength to that required to cleave the linker molecule. Suitable linkers include those based on O-nitrobenyl compounds and nitroveratryl compounds. Linkers based on benzoin

chemistry can also be used (Lee et al., J. Org. Chem. 64:3454-3460, 1999).

D. Cleavage under reductive conditions

There are many linkers known that are susceptible to reductive cleavage. Catalytic hydrogenation using palladium-based catalysts has been used to cleave benzyl and benzyloxycarbonyl groups. Disulphide bond reduction is also known in the art.

E. Cleavage under oxidative conditions

Oxidation-based approaches are well known in the art. These include oxidation of p-alkoxybenzyl groups and the oxidation of sulphur and selenium linkers. The use of aqueous iodine to cleave disulphides and other sulphur or selenium-based linkers is also within the scope of the invention.

F. Safety-catch linkers

Safety-catch linkers are those that cleave in two steps. In a preferred system the first step is the generation of a reactive nucleophilic center followed by a second step involving an intra-molecular cyclization that results in cleavage. For example, levulinic ester linkages can be treated with hydrazine or photochemistry to release an active amine, which can then be cyclised to cleave an ester elsewhere in the molecule (Burgess et al., J. Org. Chem. 62:5165-5168, 1997).

G. Cleavage by elimination mechanisms

Elimination reactions can also be used. For example, the base-catalysed elimination of groups such as Fmoc and cyanoethyl, and palladium-catalysed reductive elimination of allylic systems, can be used.

As well as the cleavage site, the linker can comprise a spacer unit. The spacer distances the nucleotide base from the cleavage site or label. The length of the linker is unimportant provided that the label is held a sufficient distance from the nucleotide so as not to interfere with any interaction between the nucleotide and an enzyme.

In a preferred embodiment the linker may consist of the same functionality as the block. This will make the deprotection and deblocking process more efficient, as only a single treatment will be required to remove both the label and the block.

A method for determining the sequence of a target polynucleotide can be carried out by contacting the target polynucleotide separately with the different nucleotides to form the complement to that of the target polynucleotide, and detecting the incorporation of the nucleotides. Such a method makes use of polymerisation, whereby a polymerase enzyme extends the complementary strand by incorporating the correct nucleotide complementary to that on the target. The polymerisation reaction also requires a specific primer to initiate polymerisation.

For each cycle, the incorporation of the modified nucleotide is carried out by the polymerase enzyme, and the incorporation event is then determined. Many different polymerase enzymes exist, and it will be evident to the person of ordinary skill which is most appropriate to use. Preferred enzymes include DNA polymerase I, the Klenow fragment, DNA polymerase III, T4 or T7 DNA polymerase, Taq polymerase or vent polymerase. A polymerase engineered to have specific properties can also be used.

The sequencing methods are preferably carried out

with the target polynucleotide arrayed on a solid support. Multiple target polynucleotides can be immobilised on the solid support through linker molecules, or can be attached to particles, e.g., microspheres, which can also be attached to a solid support material.

The polynucleotides can be attached to the solid support by a number of means, including the use of biotin-avidin interactions. Methods for immobilizing polynucleotides on a solid support are well known in the art, and include lithographic techniques and "spotting" individual polynucleotides in defined positions on a solid support. Suitable solid supports are known in the art, and include glass slides and beads, ceramic and silicon surfaces and plastic materials. The support is usually a flat surface although microscopic beads (microspheres) can also be used and can in turn be attached to another solid support by known means. The microspheres can be of any suitable size, typically in the range of from 10 nm to 100 nm in diameter. In a preferred embodiment, the polynucleotides are attached directly onto a planar surface, preferably a planar glass surface. Attachment will preferably be by means of a covalent linkage. Preferably, the arrays that are used are single molecule arrays that comprise polynucleotides in distinct optically resolvable areas, e.g., as disclosed in International App. No. WO 00/06770.

The sequencing method can be carried out on both single polynucleotide molecule and multi-polynucleotide molecule arrays, i.e., arrays of distinct individual polynucleotide molecules and arrays of distinct regions comprising multiple copies of one individual polynucleotide molecule. Single molecule arrays allow each individual polynucleotide to be resolved

separately. The use of single molecule arrays is preferred. Sequencing single molecule arrays non-destructively allows a spatially addressable array to be formed.

The method makes use of the polymerisation reaction to generate the complementary sequence of the target. The conditions necessary for polymerisation to occur will be apparent to the skilled person.

To carry out the polymerase reaction it will usually be necessary to first anneal a primer sequence to the target polynucleotide, the primer sequence being recognised by the polymerase enzyme and acting as an initiation site for the subsequent extension of the complementary strand. The primer sequence may be added as a separate component with respect to the target polynucleotide. Alternatively, the primer and the target polynucleotide may each be part of one single stranded molecule, with the primer portion forming an intramolecular duplex with a part of the target, i.e., a hairpin loop structure. This structure may be immobilised to the solid support at any point on the molecule. Other conditions necessary for carrying out the polymerase reaction, including temperature, pH, buffer compositions etc., will be apparent to those skilled in the art.

The modified nucleotides of the invention are then brought into contact with the target polynucleotide, to allow polymerisation to occur. The nucleotides may be added sequentially, i.e., separate addition of each nucleotide type (A, T, G or C), or added together. If they are added together, it is preferable for each nucleotide type to be labelled with a different label.

This polymerisation step is allowed to proceed for a time sufficient to allow incorporation of a nucleotide.

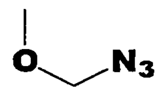
Nucleotides that are not incorporated are then removed, for example, by subjecting the array to a washing step, and detection of the incorporated labels may then be carried out.

Detection may be by conventional means, for example if the label is a fluorescent moiety, detection of an incorporated base may be carried out by using a confocal scanning microscope to scan the surface of the array with a laser, to image a fluorophore bound directly to the incorporated base. Alternatively, a sensitive 2-D detector, such as a charge-coupled detector (CCD), can be used to visualise the individual signals generated. However, other techniques such as scanning near-field optical microscopy (SNOM) are available and may be used when imaging dense arrays. For example, using SNOM, individual polynucleotides may be distinguished when separated by a distance of less than 100 nm, e.g., 10 nm to 10 µm. For a description of scanning near-field optical microscopy, see Moyer et al., Laser Focus World 29:10, 1993. Suitable apparatus used for imaging polynucleotide arrays are known and the technical set-up will be apparent to the skilled person.

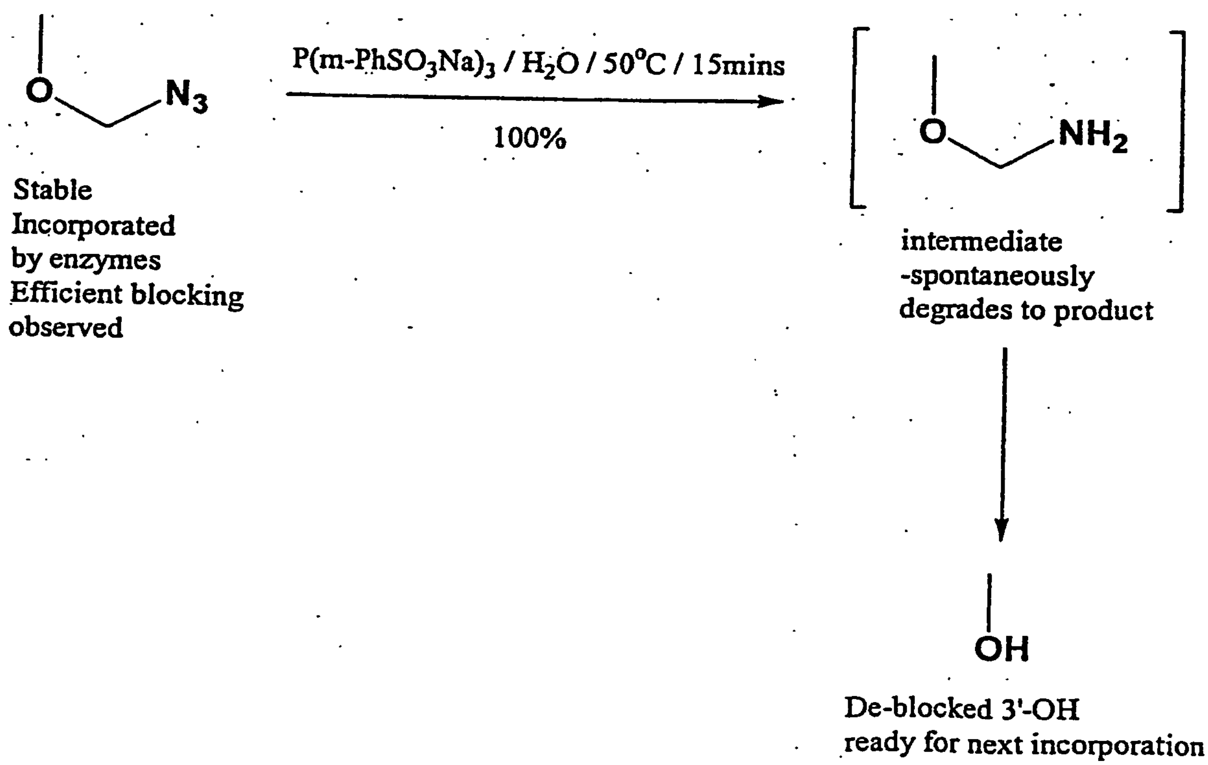
After detection, the label may be removed using suitable conditions that cleave the linker and the 3' block to allow for incorporation of further modified nucleotides of the invention.

EXAMPLES

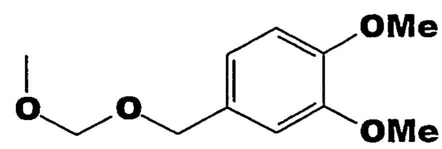
1. 3'-OH protected with an azidomethyl group as a protected form of a hemiaminal:



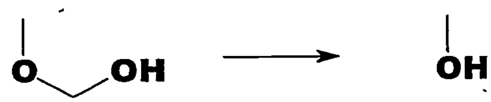
Nucleotides bearing this blocking group at the 3'position have been shown to be successfully incorporated by a number of different polymerases block efficiently and may be subsequently removed under neutral, aqueous conditions using water soluble phosphines or thiols allowing further extension:



2. 3'-OH protected with a 3,4-dimethoxybenzyloxymethyl group as a protected form of a hemiacetal:



Nucleotides bearing this blocking group have similar properties to the above example, though incorporate less rapidly. Deblocking can be achieved efficiently by the use of aqueous buffered cerium ammonium nitrate or DDQ, both conditions initially liberating the hemiacetal (1) which decomposes to the required (2) prior to further extension:



(1) (2)

The 3'-OH may also be protected with benzyl groups where the phenyl group is unsubstituted, e.g. with benzyloxymethyl, as well as benzyl groups where the phenyl group bears electron-donating substituents; an example of such an electron-rich benzylic protecting group is 3,4-dimethoxybenzyloxymethyl.

In contrast, electron-poor benzylic protecting groups, such as those in which the phenyl ring is substituted with one or more nitro groups, are less preferred since the conditions required to form the intermediate groups of formulae $-C(R')_2-OH$, $-C(R')_2-NH_2$, and $-C(R')_2-SH$ are sufficiently harsh that the integrity of the polynucleotide can be affected by the conditions needed to deprotect such electron-poor benzylic protecting groups.

3. 3'-OH protected with a fluoromethyloxymethyl group as a protected form of a hemiacetal:

$$-O-CH_2-F$$

Nucleotides bearing this blocking group may be converted to the intermediate hemiacetal using catalytic reactions known to those skilled in the art such as , for example, those using heavy metal ions such as silver.

Claims

1. A modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-OH blocking group covalently attached thereto, such that the 3' carbon atom has attached a group of the structure

$$-O-X$$

wherein X is any of $-C(R')_2-O-R$, $-C(R')_2-N(R)_2$,$-C(R')_2-N(H)R$, $-C(R')_2-S-R$ and $-C(R')_2-F$,

wherein each R group is or is part of a removable protecting group;

each R′ is independently H or an alkyl; and wherein said molecule may be reacted to yield an intermediate in which each R group is exchanged for H or, where X is

$-C(R')_2-F$, the F is exchanged for OH, which intermediate dissociates under aqueous conditions to afford a molecule with a free 3'OH;

with the proviso that where X is $-C(R')_2-S-R$, both R′ groups are not H.

2. A molecule according to claim 1 or 2, wherein R is a benzyl or substituted benzyl group.

3. A molecule according to claim 1 wherein X is an azidomethyl group.

4. A molecule according to any preceding claim wherein said base is linked to a detectable label via a cleavable linker.

5. A molecule according to claim 4, wherein said detectable label is a fluorophore.

6. A molecule according to claim 4 or 5 wherein said linker is acid labile, photolabile or contains a disulphide linkage.

7. A method of controlling the incorporation of a nucleotide molecule complementary to the nucleotide in a target single stranded polynucleotide in a synthesis or sequencing reaction comprising incorporation into the growing complementary polynucleotide a molecule according to any of claims 1 to 6 the incorporation of said molecule preventing or blocking introduction of subsequent nucleoside or nucleotide molecules into said growing complementary polynucleotide.

8. A method according to claim 7, wherein the incorporation of said molecule is accomplished by a terminal transferase or polymerase or a reverse transcriptase.

9. A method for determining the sequence of a target single stranded polynucleotide, comprising monitoring the sequential incorporation of complementary nucleotides, wherein the nucleotide is a molecule

according to any of claims 1 to 6 and include a detectable label linked to the base via a cleavable linker and wherein the identity of each nucleotide molecule incorporated is determined by detection of the label linked to the base, and said blocking group and said label is removed prior to introduction of the next complementary nucleotide.

10. A method according to claim 9, wherein said blocking group and said label are removed in a single step.

11. A method for determining the sequence of a target single-stranded polynucleotide, comprising:

(a) providing nucleotides according to any of claims 1 to 3 linked via the base to a detectable label via a cleavable linker and wherein the detectable label linked to each type of nucleotide can be distinguished upon detection from the detectable label used for other types of nucleotides;

(b) incorporating the nucleotide into the complement of the target single stranded polynucleotide;

(c) detecting the label of the nucleotide of (b), thereby determining the type of nucleotide incorporated;

(d) removing the label of the nucleotide of (b) and the blocking group; and

(e) optionally repeating steps (b)-(d) one or more times;

thereby determining the sequence of a target single-stranded polynucleotide.

12. A method according to claim 11 wherein the label of the nucleotide and the blocking group are removed in a single chemical treatment step.

13. A method according to claim 12 or 13, wherein each of the nucleotides are brought into contact with the target sequentially, with removal of non-incorporated nucleotides prior to addition of the next nucleotide, and wherein detection and removal of the label and the blocking group is carried out either after addition of each nucleotide, or after addition of all four nucleotides.

14. A method according to claim 12 or 13, wherein each of the nucleotides are brought into contact with the target together simultaneously, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and the blocking group.

15. A method according to claim 12 or 13, comprising a first step and a second step, wherein in the first step, a first composition comprising two of the four nucleotides is brought into contact with the target and non-incorporated nucleotides are removed prior to detection and subsquent to removal of the label, and wherein in the second step, a second composition comprising the two nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group, and wherein the first and second steps are optionally repeated one or more times.

16. A method according to claim 12 or 13 comprising a first step and a second step, wherein in the first step, a composition comprising one of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein in the second step, a second composition comprising the three nucleotides not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein the first steps and the second step are optionally repeated one or more times.

17. A method according to claim 15 comprising a first step and a second step, wherein in the first step, a first composition comprising three of the four nucleotides is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein in the second step, a composition comprising the nucleotide not included in the first composition is brought into contact with the target, and non-incorporated nucleotides are removed prior to detection and subsequent to removal of the label and blocking group and wherein the first steps and the second step are optionally repeated one or more times.

18. A kit, comprising:

(a) individual nucleotides according to any of claims 1 to 3; and

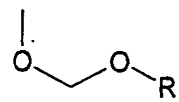
(b) packaging materials therefor.

19. A kit according to claim 18, wherein said nucleotides are linked via the base to a detectable label by a cleavable linker, and wherein the detectable label can be distinguished upon detection from the detectable label used for any of the other three types of nucleotide.

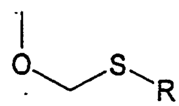
20. The kit of claim 18 or 19, further comprising an enzyme and buffers appropriate for the action of the enzyme.

: 416307: CDM: MMC: LONDOCS

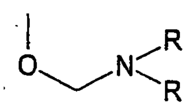
Protected hemiacetals:



Protected hemithioacetals:



Protected hemiaminals:



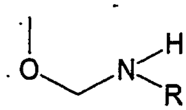
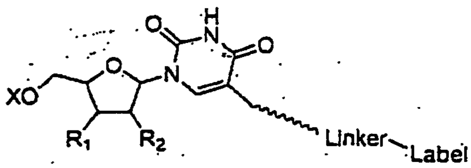
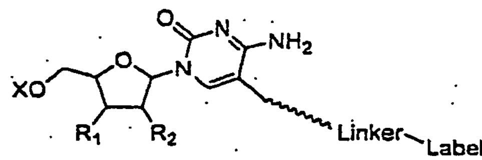


Fig. 1

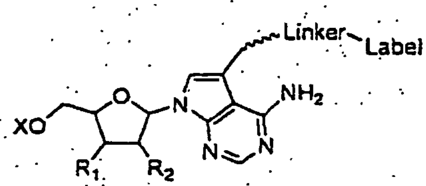
THIS PAGE



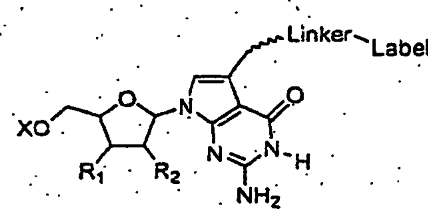
Uridine C5-linker



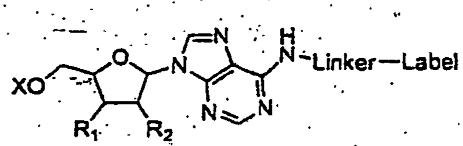
Cytidine C5-linker



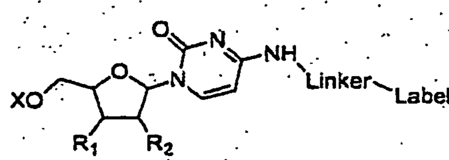
N7 Deazaadenosine C7-linker



N7 Deazaguanosine C7-linker



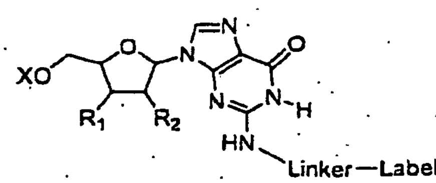
Adenosine N6-linker



Cytidine N4-linker

where R_1 and R_2, which may be the same or different, are each selected from H, OH, or any group which can be transformed into an OH. Suitable groups for R_1 and R_2 are described in Figure 4

X = H, phosphate, diphosphate or triphosphate



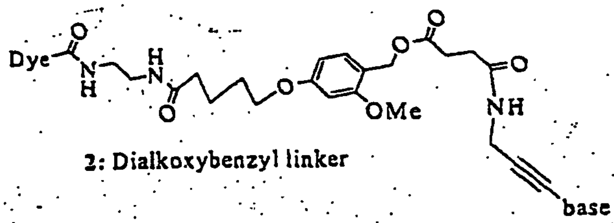
Guanosine N2-linker

Fig. 2

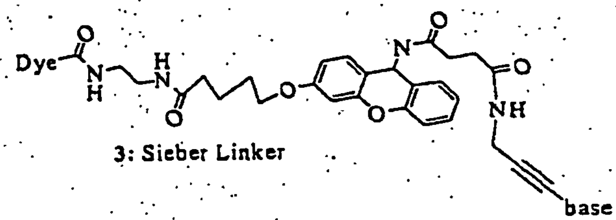
Fig. 2

THIS PAGE IS

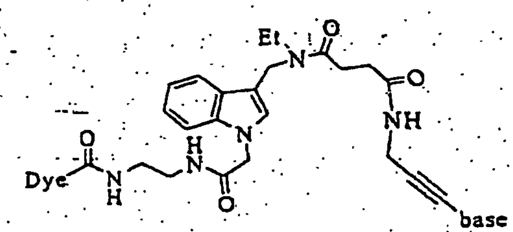
Acid Labile Linkers



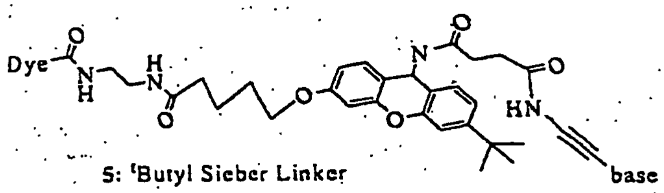
2: Dialkoxybenzyl linker



3: Sieber Linker

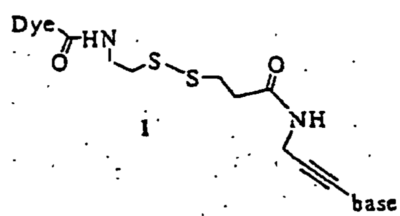


4: Indole Linker



5: ᵗButyl Sieber Linker

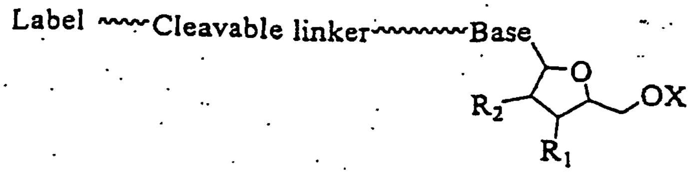
Disulfide Linker



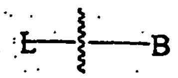
1

Fig. 3

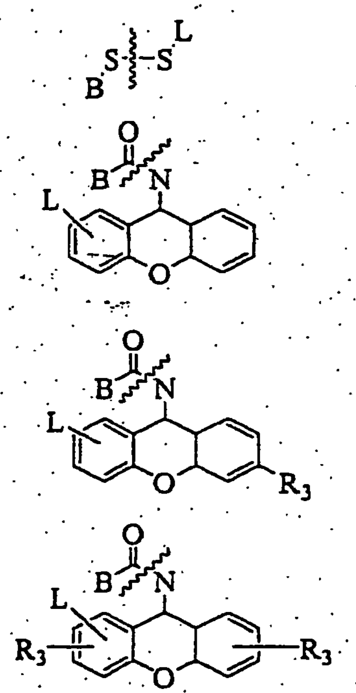
THIS PAGE IS BLANK



where R_1 and R_2, which may be the same or different, are each selected from H, OH, or any group which can be transformed into an OH, including a carbonyl



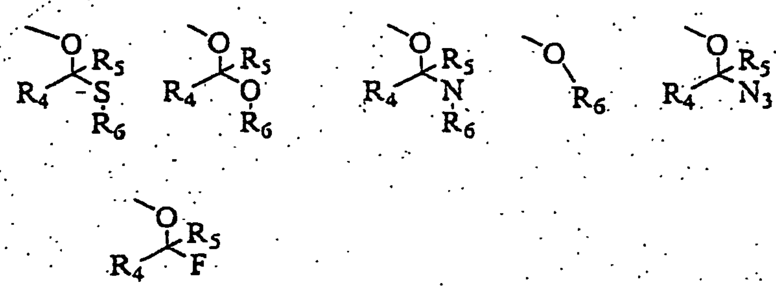
Cleavable linkers may include:



R3 represents one or more substituents independently selected from alkyl, alkoxy, amino or halogen

Alternatively, cleavable linkers may be constructed from any labile functionality used on the 3'-block

R_1 and R_2 groups may include



where R_4 is H or alkyl, R_5 is H or alkyl and R_6 is alkyl, cycloalkyl, alkenyl, cycloalkenyl or benzyl

and X is H, phosphate, diphosphate or triphosphate

Fig. 4